SOME RELATIONS BETWEEN SIGNAL DETECTION AND THE CAPACITY OF COMMUNICATION CHANNELS*

C.R. Baker
Department of Statistics
University of North Carolina
Chapel Hill, N.C. 27599-3260

LISS-27

October, 1988

*Presented at the Conference on Advances in Communications and Control (ComCon 88), Baton Rouge, Louisiana, Oct. 19-21, 1988.

The research for this paper was supported by the Office of Naval Research (Contract N00014-86-K-0039) and by the National Science Foundation (Grant NCR-8713726).

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for Public Release: Distribution Unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Department of Statistics 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State and ZIP Code) University of North Carolina Chapel Hill, North Carolina 27514	7b. ADDRESS (City, State and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-86-K-0039
8c. ADDRESS (City, State and ZIP Code) Statistics & Probability Program Arlington, VA 22217	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT NO.
11. TITLE (Include Security Classification) Some Relations Between Signal Detection (cont.)	
12. PERSONAL AUTHOR(S) C.R. Baker	
13a. TYPE OF REPORT TECHNICAL 13b. TIME COVERED FROM ______ TO ______	14. DATE OF REPORT (Yr., Mo., Day) October, 1988 15. PAGE COUNT 12
16. SUPPLEMENTARY NOTATION	
17. COSATI CODES FIELD / GROUP / SUB. GR.	18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Information theory; signal detection; channel capacity.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Calculation of the capacity of information channels is shown to be closely related to some basic results on detection of signals in noise.

11. TITLE CONT. And the Capacity of Communication Channels.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED XX SAME AS RPT. XX DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION
22a. NAME OF RESPONSIBLE INDIVIDUAL C.R. Baker	22b. TELEPHONE NUMBER (Include Area Code) (919) 962-2189 22c. OFFICE SYMBOL

DD FORM 1473, 83 APR   EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

0. Introduction.

The theme of the ComCon conferences is the unification of communications and control. Thus, methods and results that are common to the two areas are of central interest. This paper contains a discussion of a topic that is in this general spirit: relations that exist between signal detection and information theory. These relations are not contained in existing textbooks on information theory and/or signal detection. They have been developed in recent years as a result of specific needs: they are essential in obtaining solutions to channel capacity problems when the noise sample paths comprise an infinite-dimensional linear manifold.

Channel capacity is one of the most basic problems of information theory. In many setups, such as that of the DMC (discrete memoryless channel), the basic mathematical structure is so simple that measure-theoretic questions do not arise. The situation changes radically when one considers more complicated channels, such as the continuous-time Gaussian channel with memory. Even in this case, the actual complexity of the problem is sometimes masked by mathematically inadmissible simplifications and/or by simply ignoring some of the more difficult problems. In fact, a rigorous mathematical analysis of capacities of such channels, which include some of the most widely-used models in information theory, cannot be completely carried out without the use of measure theory. While this may be widely recognized, it may be less obvious that there are some very strong relations between channel capacity problems and some well-known problems in signal detection.

Before proceeding, it is necessary to define what will be meant here by signal detection. The problems to be considered are parametric: the statistical distributions are known for both the noise process and the signal-and-noise process. One of the first questions that need to be considered in such a problem is whether or not the mathematical model is reasonable. Since most practical problems do not permit zero probability of error in deciding whether or not signal is present (singular detection), the mathematical model should first be checked to determine if the likelihood ratio (Radon-Nikodym derivative) exists. If it does exist, then various criteria dictate the use of the likelihood ratio, or some monotone function of the likelihood ratio, as the test statistic. For the purposes of this paper, "signal detection" is limited to these two aspects: conditions for existence of the likelihood ratio, and (when such conditions are satisfied) the form of the likelihood ratio. These are important to any detection problem. Not considered here are more specialized aspects of signal detection, such as error probabilities, approximation of optimum detectors, etc. It will be seen that existence of, and expressions for, likelihood ratios are essential to some of the basic results in channel capacity.

The following sections give some preliminary definitions, then a discussion of relations between information capacity and signal detection, followed by a discussion of relations between coding capacity and signal detection.

1. Basic Assumptions and Definitions.

The problems to be considered are those where the processes of interest have sample functions belonging to a real separable Hilbert space, H, with inner product $\langle\cdot,\cdot\rangle$ and associated norm $\|\cdot\|$. The noise process will always be denoted by N; it corresponds to a probability measure μ_N. It will be assumed WLOG that H is the smallest closed set A such that $\mu_N[A] = 1$. All measures will be probability measures; they are always understood to be defined on the

Borel σ-field, denoted by $\mathbb{B}[H]$. In particular, measures on $H \times H$ are on the σ-field $\mathbb{B}[H] \times \mathbb{B}[H]$. If μ_Z and μ_W are two measures on H (or two on $H \times H$), then $\mu_Z << \mu_W$ means that μ_Z is absolutely continuous with respect to μ_W: for any measurable set A, $\mu_W(A) = 0$ implies that $\mu_Z(A) = 0$. If the measures are mutually absolutely continuous, this will be denoted by $\mu_Z \sim \mu_W$. In the language of signal detection, the detection problem is then nonsingular. Singular detection is understood to mean a problem where zero probability of error is possible in making decisions. Usually nonsingular detection is taken to mean that mutual absolute continuity holds, and it will be seen that this is the important property in applications to channel capacity. The situation where the measures are neither orthogonal nor mutually absolutely continuous is sometimes called "intermediate-singular"; it arises less frequently in the channel capacity problems.

To each such noise process N, there exists a covariance operator R_N; it is defined by

$$\langle R_N u, v \rangle = \int_H \langle x, u \rangle \langle x, v \rangle d\mu_N(x).$$

The process N is assumed to have zero mean. R_N is self-adjoint, strictly positive, and will be assumed to have finite trace. The latter condition is satisfied, for example, when $H = L_2[0,T]$ and $N = (N_t)$ is a m.s. continuous stochastic process. In that case, R_N is simply the integral operator having the noise covariance function as its kernel.

The communication channel to be considered is defined as follows. If X is the channel input (either a sample function from a stochastic process or a member of a set of code words), then the channel output Y is defined as $Y = f(X,N)$, where $f\colon H \times H \to H$ is a measurable mapping. In the case that X is a stochastic process, it will be assumed that X and N are statistically independent, so that their joint probability measure μ_{XN} on H×H satisfies $\mu_{XN} = \mu_X \otimes \mu_N$. $\otimes$ denotes product measure. In the additive channel, $f(x,y) = A(x) + y$, where A is some coding or constraint function. However, in the interest of generality, the channel will not initially be assumed to be additive.

Given these definitions, the known signal and stochastic signal detection problems can be identified. The known signal detection problem is for testing the hypothesis that an observation is from the process N against the hypothesis that the observation is from the process f(x,N), where x is fixed. The two measures of interest are then μ_N and $\mu_N \circ f_x^{-1}$, where $f_x(y) = f(x,y)$. Thus, $\mu_N \circ f_x^{-1}[A] = \mu_N\{y: f(x,y) \in A\}$. The stochastic signal detection problem involves the same noise process, but now the signal-and-noise process is $Y = f(X,N)$, where X is a stochastic process:

$$\mu_Y(A) = \mu_X \otimes \mu_N \{(x,y): f(x,y) \in A\}.$$

In the case of simple channels, such as those without memory, information capacity and coding capacity are typically equal. This may no longer be true for more general channels. In general, due to Fano's inequality [1], information capacity suitably defined is an upper bound on coding capacity. In the following, relations to signal detection are discussed for both of these definitions of channel capacity.

The probabilities of interest have been defined as being on the Borel σ-field of a Hilbert space. However, the structure can be more general. Some of the results to be discussed do not require a linear space structure. Others can be extended to more general linear spaces, such as a class of linear topological vector spaces [2].

2. Mutual Information, Information Capacity, and Signal Detection.

Suppose that the channel input is a sample function from a stochastic process X. Based on Shannon's original definition, the (average) mutual information between X and the channel output Y can be expressed as follows. Let μ_{XY} be the joint probability measure of (X,Y) on H×H; it is defined by

$$\mu_{XY}(A) = \mu_X \otimes \mu_N \{(x,y): (x, f(x,y)) \in A\}.$$

The general definition for mutual information I(X,Y) is then defined to be

$$\sup_{N;A_1,\dots,A_N} \sum_{i=1}^{N} \mu_{XY}(A_i) \log \frac{\mu_{XY}(A_i)}{\mu_X \otimes \mu_Y(A_i)}$$

In this expression, the supremum is taken over all $N \geq 1$ and all measurable partitions $A_1,\dots,A_N$ of the product space H×H. Such a definition is of course very widely used when the probabilities of interest are for discrete random variables. An expression that is more convenient for general probabilities has been proved by Dobrushin [3]. That is, the mutual information of X and Y is infinite if it is false that $\mu_{XY} << \mu_X \otimes \mu_Y$; otherwise, its value (which may be infinite) is given by

$$I(X,Y) = \int_{H\times H} \log[d\mu_{XY}/d\mu_X \otimes \mu_Y]d\mu_{XY}.$$

It is clear that this definition, one of the most basic for information theory, involves absolute continuity and calculation of a likelihood ratio. For a particular channel (a triplet [f,X,N]), one would like to have a general method of determining if absolute continuity of μ_{XY} and $\mu_X \otimes \mu_Y$ holds, and -- if it does hold -- finding the likelihood ratio $d\mu_{XY}/d\mu_X \otimes \mu_Y$. In fact, such methods can be applied, based on the known signal and stochastic signal detection problems.

<u>Theorem 1</u> [4]: If the known signal detection problem is nonsingular for almost all x:

$$\mu_N \sim \mu_N \circ f_x^{-1} \quad \text{a.e. } d\mu_X(x),$$

then $\mu_Y \sim \mu_N$ and $\mu_{XY} \sim \mu_X \otimes \mu_Y$.

If the channel model is such that the likelihood ratio $d\mu_{XY}/d\mu_X \otimes \mu_Y$ exists, then the next order of business is to give its form. This can also be done using the known signal and stochastic signal likelihood ratios. Unfortunately, the precise statement of this condition involves extension of the original measures to completed σ-fields, as in the following theorem.

<u>Theorem 2</u> [5]: Let $\overline{\mathbb{B}[H]}$ be the completion of $\mathbb{B}[H]$ with respect to μ_X and $\overline{\mathbb{B}[H] \times \mathbb{B}[H]}$ the completion of $\mathbb{B}[H] \times \mathbb{B}[H]$ with respect to $\mu_X \otimes \mu_N$, with $\bar{\mu}_X$ and $\overline{\mu_X \otimes \mu_N}$ the extended measures. Suppose that

(a) $\mu_N \circ f_x^{-1} \sim \mu_N$ a.e. $d\mu_X(x)$;

(b) the map g: $(x,y) \to (d\mu_N \circ f_x^{-1})/d\mu_N(y)$ is $\overline{\mathbb{B}[H] \times \mathbb{B}[H]}/\mathbb{B}[R^1]$ measurable;

(c) $\int_H [\log(d\mu_Y/d\mu_N)](y) d\mu_Y(y) < \infty$.

Then

$$I(\mu_{XY}) = \int_H \int_H [\log(d\mu_N \circ f_x^{-1}/d\mu_N)(y)] d\mu_N \circ f_x^{-1}(y) d\bar{\mu}_X(x)$$
$$- \int_H [\log(d\mu_Y/d\mu_N)(y)] d\mu_Y(y).$$

In some important applications, condition (a) of Theorem 2 implies conditions (b) and (c), at least for some version of the function g. Thus, in such instances, the determination of the existence of the Radon-Nikodym derivative $d\mu_{XY}/d\mu_X \otimes \mu_Y$ and the calculation of the mutual information can be carried out if the known signal detection problem μ_N vs. $\mu_N \circ f_x^{-1}$ is nonsingular for almost all signal paths x, and the likelihood ratios $d\mu_N \circ f_x^{-1}/d\mu_N$ and $d\mu_Y/d\mu_N$ can be determined. A necessary condition for the known signal detection problem to be nonsingular for fixed x is that x be in the range of $R_N^{1/2}$, the square root of the covariance operator of N; this condition is also sufficient if N is Gaussian [6].

Information capacity is the supremum of the mutual information I(X,Y) over some class of admissible signal processes X. The preceding results provide insight on the constraints that should be applied if the capacity is to be finite. In the case that the channel is additive, $f(x,y) = A(x) + y$, and N is Gaussian, one can show that the conditions of Theorem 2 are satisfied when $\mu_N \sim \mu_N \circ f_x^{-1}$ for almost all x, and that an upper bound on the mutual information is given by Trace $R_N^{-1/2} R_{AX} R_N^{-1/2}$, where R_{AX} is the covariance operator of $\mu_X \circ A^{-1}$. In fact, if the channel is additive and N is Gaussian, one can show [7] that a necessary and sufficient condition for finite capacity is that each admissible signal process X and coding function A satisfy a constraint of the form $E_{\mu_X} \|A(x)\|_N^2 \leq P$ for some finite P, where $\|u\|_N^2 = \|R_N^{-1/2} u\|^2$. Now, given a general additive channel with second-order noise process N (i.e., $E_{\mu_N} \|x\|^2 < \infty$), and a constraint given in terms of the noise covariance, it is known [8] that the channel capacity is minimized if the noise is Gaussian. Thus, one can conclude that the above necessary and sufficient condition for finite capacity when the channel is additive and the noise is Gaussian is also necessary when the noise is permitted to be nonGaussian, but with trace-class covariance. Of course, the constraint may not be given in this form, but finite capacity implies existence of such a constraint. Typically, a constraint may have the form $E_{\mu_X} \|x\|_W^2 \leq P$, where $\|\cdot\|_W$ is the RKHS norm of the covariance R_0, with parameter set H, $R_0(u,v) = \langle R_W u, v \rangle$, and R_W is a self-adjoint and non-negative bounded linear operator in H. If $R_W > 0$, then

$\|x\|_W^2 = \|R_W^{-1/2}x\|^2$. If $R_W \neq R_N$, such a constraint results in a "mismatched" channel, meaning that the constraint covariance is not matched to the channel noise covariance. One can consider R_W to be the covariance operator of a zero mean Gaussian process W. One will wish to consider relations that must exist between N and W in order that the capacity be finite, and here again the result has striking similarities to well-known results in signal detection.

A fundamental result for the stochastic signal detection problem when both processes are zero-mean Gaussian is the following: mutual absolute continuity of μ_N and μ_W holds if and only if the two covariance operators satisfy $R_N = R_W^{1/2}(I+S)R_W^{1/2}$ where S is Hilbert-Schmidt and $(I+S)^{-1}$ exists [6]. This can be compared with the following condition for finite information capacity.

Theorem 3 [9]: Suppose that N is Gaussian, that $f(x,y) = A(x) + y$, and that X and A must satisfy the constraint $E_{\mu_X}\|A(x)\|_W^2 \leq P$. Then the capacity will be finite if and only if there exists a densely-defined self-adjoint operator S in H such that $(I+S)^{-1}$ exists and is bounded, and $R_N = R_W^{1/2}(I+S)R_W^{1/2}$.

It follows that the information capacity of the mismatched channel will be finite if the constraint RKHS norm $\|\cdot\|_W$ is given by a covariance operator R_W corresponding to a zero-mean Gaussian measure μ_W such that $\mu_W \sim \mu_N$. However, this sufficient condition is obviously not a necessary condition.

3. Coding Capacity and Signal Detection.

Coding capacity requires a more specific discussion than information capacity, since the definition of a code depends on the nature of the channel. For simplicity, only the additive channel will be considered here, and for the continuous-time channel: the code words must be elements of $L_2[0,T]$, where T is permitted to become large.

Let T be fixed. A code (k, F_T, ϵ) is a set of k elements of $L_2[0,T]$, say $x_1, x_2, \ldots, x_k$, each belonging to the set $F_T \in B[L_2[0,T]]$, together with a measurable partition $C_1, \ldots, C_k$ of $L_2[0,T]$ such that $\mu_N^T \circ f_{x_i}^{-1}[C_i] \geq 1-\epsilon$ for $i = 1,2,\ldots,k$. μ_N^T denotes the probability measure defined by N when the parameter set is restricted to $[0,T]$. Typically, F_T will be defined as the set of all elements x satisfying a constraint of the form $\|x\|_{W,T}^2 \leq PT$, where $\|\cdot\|_{W,T}$ is a RKHS norm, defined by a fixed covariance function r_W. A non-negative real number R is said to be a permissible rate of transmission if there exists a sequence of codes $([e^{T_i R}], T_i, \epsilon_i)$ such that $T_i \to \infty$ and $\epsilon_i \to 0$ as $i \to \infty$; $[r]$ is the integer part of r. The channel coding capacity is then the supremum over the set of all admissible R.

The additive channel noise has a covariance operator $R_{N,T}$ with finite trace, for each $T > 0$. Thus, there exists a zero-mean Gaussian process G with covariance operator $R_{G,T} = R_{N,T}$. For any fixed T, the relative entropy of N with respect to G, whenever $\mu_N^T << \mu_G^T$, is

$$H_G^T(N) = \int_H [\log d\mu_N^T / d\mu_G^T] d\mu_N^T.$$

If $\mu_N^T << \mu_G^T$ is false, then the relative entropy is defined to be infinite. Two basic tools for determining coding capacity are Fano's inequality [1] and Feinstein's Lemma (as modified and generalized by Thomasian, Kadota, and McKeague). From Fano's inequality, one can obtain the following result.

Theorem 4: An upper bound on the coding capacity is given by

$$C \leq \overline{\lim_{T\to\infty}} \left[\frac{1}{T} C_W^T(P) + H_G^T(N)\right]$$

where $C_W^T(P)$ is the information capacity of the Gaussian channel previously discussed when $H = L_2[0,T]$ and the constraint is given by $E_{\mu_X} \|x\|_{W,T}^2 \leq PT$.

As can be seen, the upper bound obtained from Fano's inequality depends heavily on results concerning absolute continuity.

The lower bound on the coding capacity can be determined from Feinstein's Lemma, often called the Fundamental Lemma. It can be stated as follows:

Theorem 5 [10]: Suppose that $\mu_N \circ f_x^{-1} \sim \mu_N$ a.e. $d\mu_X(x)$, and that the map g: $(x,y) \to [d\mu_N \circ f_x^{-1}/d\mu_Y](y)$ is $\mathbb{B}[H]\times\mathbb{B}[H]/\mathbb{B}[\mathbb{R}^1]$ measurable. For any real number α let $A = \{(x,y) \in H\times H: \log[d\mu_{XY}/d\mu_X \otimes \mu_Y](x,y) > \alpha\}$. Then for each positive integer k and $F \in \mathbb{B}[H]$ there exists a code (k,F,ϵ) such that

$$\epsilon \leq ke^{-\alpha} + \mu_{XY}(A^c) + \mu_X(F^c).$$

In applying Feinstein's Lemma, the set F is typically taken to be those x satisfying a constraint. Thus, for a code (k,F_T,ϵ) as defined above for the time-continuous additive channel, the measures appearing in the theorem are on $L_2[0,T]$ or $L_2[0,T]\times L_2[0,T]$, and F would be the set of those x in $L_2[0,T]$ such that $\|x\|^2_{W,T} \leq PT$. The application of Feinstein's Lemma is to show that $\mu_X^T(F_T^c) \to 0$ and $\mu_{XY}^T(A_T^c) \to 0$, as $T \to \infty$ along a subsequence. The constraint $\alpha = \alpha(T)$ is then related to the constraint and the channel noise.

It can be seen that Feinstein's Lemma, in the complex setup of the continuous-time channel, supposes that the known signal problem be nonsingular. When this occurs, then from above, $\mu_{XY} \sim \mu_X \otimes \mu_Y$ and in some important channels the Radon-Nikodym derivative $d\mu_{XY}/d\mu_X \otimes \mu_Y$ can be computed using the known signal and stochastic signal likelihood ratios.

4. Channels with Feedback.

Although the previous results can be applied for channels with feedback, the introduction of feedback substantially complicates the capacity problem. In fact, at present the value of the capacity is not known, for either the information capacity or the coding capacity, for any mismatched channel with memory where the noise covariance R_N has infinite-dimensional range space. This includes not only the continuous-time channels, but also the simplest of all Gaussian channels with memory: the additive discrete-time channel.

Analysis of feedback channels for continuous-time channels is more conveniently carried out using a stochastic calculus formulation. For an illustration of the relations that exist between absolute continuity, likelihood ratio, and channel capacity in this framework, see [11].

5. Concluding Remarks.

This discussion has described some relations that exist between well-known signal detection problems and recent results that have been instrumental in determining capacity of communication channels with memory. Much work remains to be done for such channels, and it can be expected that these relations will continue to play an important role in modeling, analysis, and solution of the many open problems.

References.

[1] R.B. Ash, *Information Theory*, Interscience, New York; 1965.

[2] C.R. Baker, Information capacity of Gaussian channels, in *Recent advances in Communication and Control Theory*, R.E. Kalman, G.I. Marchuk, A.E. Ruberti, and A.J. Viterbi, Eds., Optimization Software, Inc., Publ. Div., 2-15, 1987.

[3] R.L. Dobrushin, General formulation of Shannon's basic theorems of the theory of information, *Uspehi Mat. Neuk.* *14* , 3-104, 1959; translated in *Amer. Math.* Soc. *Transl.* (2) *12*, 199-246, 1959.

[4] C.R. Baker, Absolute continuity and applications to information theory, in *Lectures Notes in Mathematics No.* 526, 1-11, Springer-Verlag, Berlin/New York, 1976.

[5] C.R. Baker, Calculation of the Shannon information, *J. Math. Anal. & Appl.* 69, 115-123, 1979.

[6] C.R. Rao and V.S. Varadarajan, Discrimination of Gaussian processes, *Sankhyā Ser. A* 25, 303-330, 1963.

[7] C.R. Baker, Capacity of the Gaussian channel without feedback, *Inform. Contr.* 37, 70-89, 1978.

[8] S. Ihara, On the capacity of channels with additive nonGaussian noise, *Inform. Contr.* 37, 34-39, 1978.

[9] C.R. Baker, Capacity of the mismatched Gaussian channel, *IEEE Trans. Inform. Th.* IT-33, No. 6, 802-812, 1987.

[10] I.W. McKeague, On the capacity of channels with Gaussian and nonGaussian noise, *Inform. Contr.* 51, 153-173, 1981.

[11] C.R. Baker and A.F. Gualtierotti, Signal detection and channel capacity for spherically-invariant processes, *Proc. 23rd Conf. on Dec. & Control*, Las Vegas, Nev., Dec. 1984, 1444-1446.